```
####

# set wd
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

getwd()

# clean workspace:
graphics.off()
rm(list=ls())

# load the function for m shapiro test:
load("C:/Users/Marco Paggiaro/OneDrive/Desktop/mieidati/Applied Statistics/
mcshapiro.test.RData")


# Import all the libraries, just because
library(class)
library(MASS)
library(car)

# data import:
stoneflakes <- read.table("stoneflakes.txt", header = T)
head(stoneflakes)
plot(stoneflakes)

# a) Clustering, Ward:
d <- dist(stoneflakes, method='euclidean')
ward.stones <- hclust(d, method='ward.D2')
plot(ward.stones, main='euclidean-ward', hang=-0.1,
xlab='',,ylab='distance', labels=F, cex=0.6, sub='')

# b)
cluster <- cutree(ward.stones, k = 3)
cluster <- as.factor(cluster)

levels(cluster)
# assumptions:
# 1) Normality of each group:
mcshapiro.test(stoneflakes[which(cluster == "1"),])
mcshapiro.test(stoneflakes[which(cluster == "2"),])
mcshapiro.test(stoneflakes[which(cluster == "3"),])

# OK!!
# 2) Same covariances:
S1 <-  cov(stoneflakes[which(cluster == "1"),])
S2 <-  cov(stoneflakes[which(cluster == "2"),])
S3 <-  cov(stoneflakes[which(cluster == "3"),])
par(mfrow=c(1,3))
image(S1, col=heat.colors(100),main='Cov. S1', asp=1, axes = FALSE, breaks
= quantile(rbind(S1,S2,S3), (0:100)/100, na.rm=TRUE))
image(S2, col=heat.colors(100),main='Cov. S2', asp=1, axes = FALSE, breaks
= quantile(rbind(S1,S2,S3), (0:100)/100, na.rm=TRUE))
image(S3, col=heat.colors(100),main='Cov. S3', asp=1, axes = FALSE, breaks
= quantile(rbind(S1,S2,S3), (0:100)/100, na.rm=TRUE))
# OK

# test:
fm <- manova(as.matrix(stoneflakes) ~ cluster )
summary(fm, test = "Wilks")
```

```
# Yes! significance.

# c)
i1 <- which(cluster == "1")
i2 <- which(cluster == "2")
i3 <- which(cluster == "3")


n1 <- length(i1)
n2 <- length(i2)
n3 <- length(i3)
n  <- n1+n2+n3


g  <- 3
p  <- 2



alpha <- 0.1
k <- p*g*(g-1)/2
qT <- qt(1-alpha/(2*k), n-g)

W <- summary.manova(fm)$SS$Residuals
m  <- sapply(stoneflakes,mean)          # estimates mu
m1 <- sapply(stoneflakes[i1,],mean)     # estimates mu.1=mu+tau.1
m2 <- sapply(stoneflakes[i2,],mean)     # estimates mu.2=mu+tau.2
m3 <- sapply(stoneflakes[i3,],mean)     # estimates mu.3=mu+tau.3

inf12 <- m1-m2 - qT * sqrt( diag(W)/(n-g)  *  (1/n1+1/n2) )
sup12 <- m1-m2 + qT * sqrt( diag(W)/(n-g)  *  (1/n1+1/n2) )
inf13 <- m1-m3 - qT * sqrt( diag(W)/(n-g)  *  (1/n1+1/n3) )
sup13 <- m1-m3 + qT * sqrt( diag(W)/(n-g)  *  (1/n1+1/n3) )
inf23 <- m2-m3 - qT * sqrt( diag(W)/(n-g)  *  (1/n2+1/n3) )
sup23 <- m2-m3 + qT * sqrt( diag(W)/(n-g)  *  (1/n2+1/n3) )

CI <- list(diff1_2=cbind(inf12, sup12), diff1_3=cbind(inf13, sup13),
diff2_3=cbind(inf23, sup23))
CI
```